HW1 - Time series graphics

CODE

Rachel Greenlee

2/4/2022

Exercises from Section 2 of Forecasting: Principles & Practice at https://otexts.com/fpp3/graphics-exercises.html (https://otexts.com/fpp3/graphics-exercises.html)

Exercise 2.1

Use the help function to explore what the series gafa_stock, PBS, vic_elec and pelt represent.

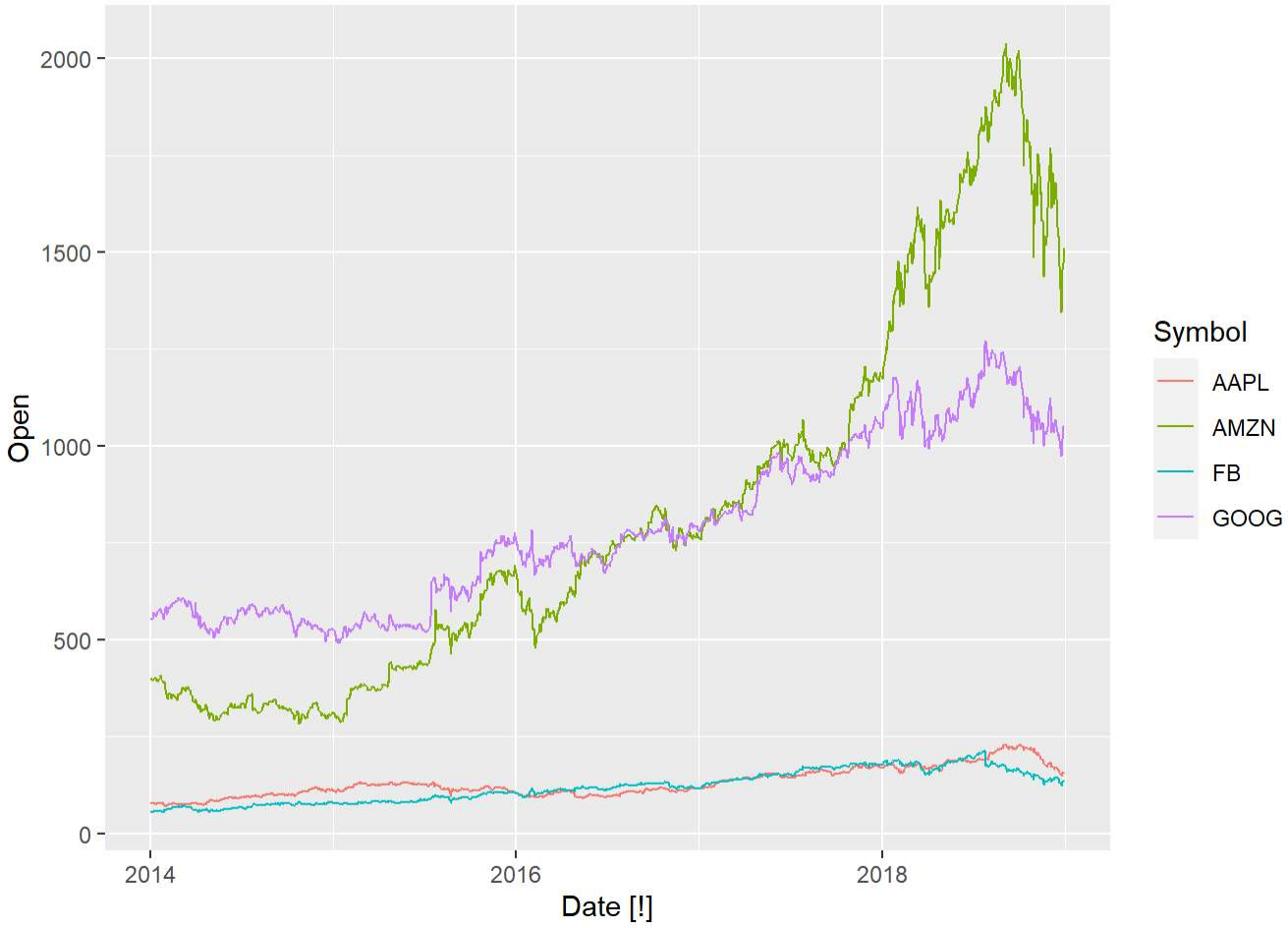
HIDE

```
#?gafa_stock
#?PBS
#?vic_elec
#?pelt
```

Use autoplot() to plot some of the series in these data sets.

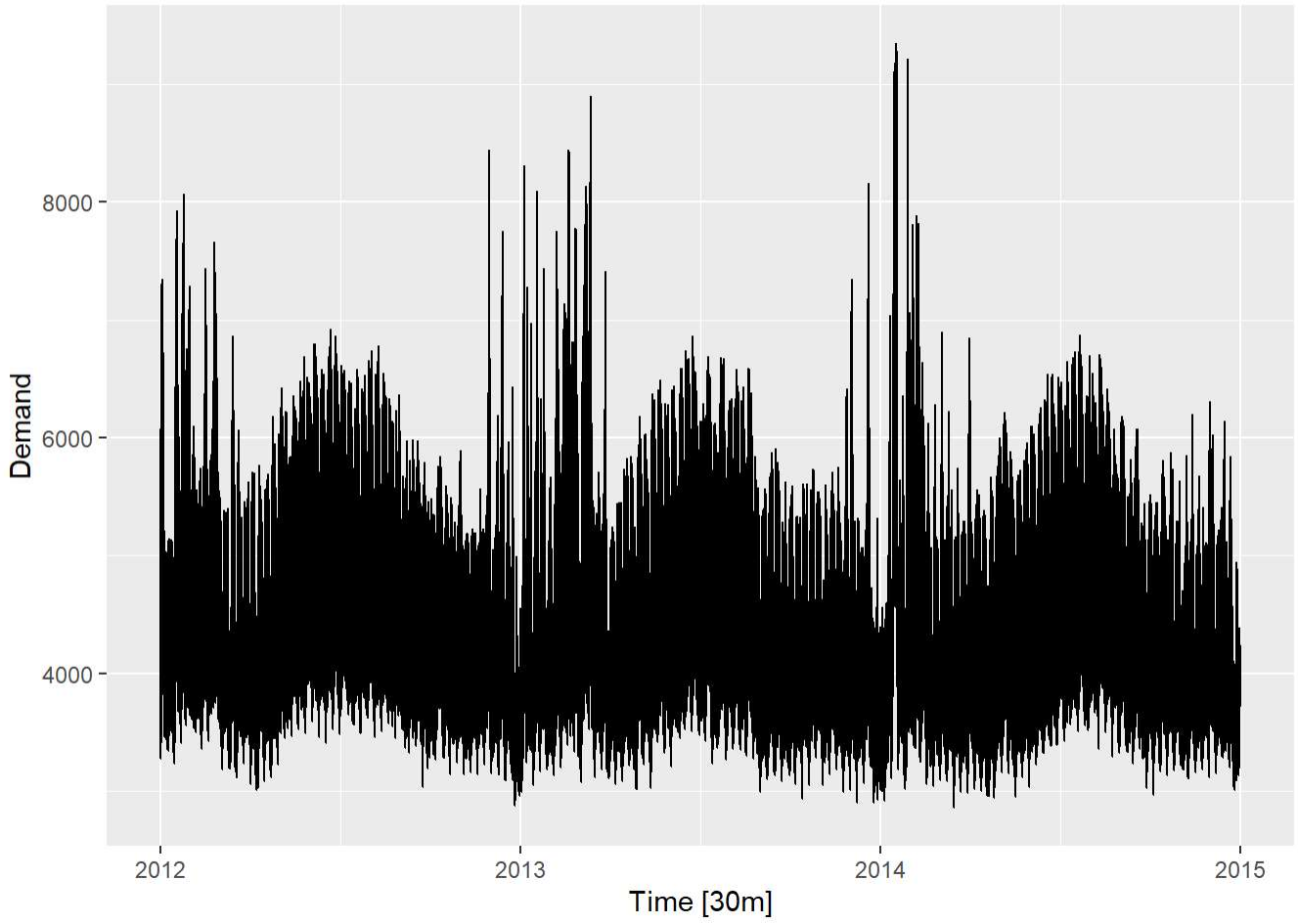
HIDE

```
autoplot(gafa_stock)
```



HIDE

```
autoplot(vic_elec)
```



What is the time interval of each series?

`gafa_stock` is daily, with the omission of weekends and federal holidays
`PBS` is monthly
`vic_elec` is in half-hour (30 minute) increments
`pelt` is annual

Exercise 2.2

Use filter() to find what days corresponded to the peak closing price for each of the four stocks in gafa_stock.

HIDE

```
gafa_stock %>%
  group_by(Symbol) %>%
  filter(Close == max(Close))
```

```
## # A tsibble: 4 x 8 [!]
## # Key:       Symbol [4]
## # Groups:    Symbol [4]
##   Symbol Date        Open  High   Low Close Adj_Close   Volume
##   <chr>  <date>     <dbl> <dbl> <dbl> <dbl>     <dbl>    <dbl>
## 1 AAPL   2018-10-03  230.  233.  230.  232.      230. 28654800
## 2 AMZN   2018-09-04 2026. 2050. 2013  2040.     2040.  5721100
## 3 FB     2018-07-25  216.  219.  214.  218.      218. 58954200
## 4 GOOG   2018-07-26 1251  1270. 1249. 1268.     1268.  2405600
```

Exercise 2.3

Download the file tute1.csv from the book website, open it in Excel (or some other spreadsheet application), and review its contents. You should find four columns of information. Columns B through D each contain a quarterly series, labelled Sales, AdBudget and GDP. Sales contains the quarterly sales for a small company over the period 1981-2005. AdBudget is the advertising budget and GDP is the gross domestic product. All series have been adjusted for inflation.

You can read the data into R with the following script:

HIDE

```
tute1 <- readr::read_csv("tute1.csv")
#View(tute1)
```

Convert the data to time series

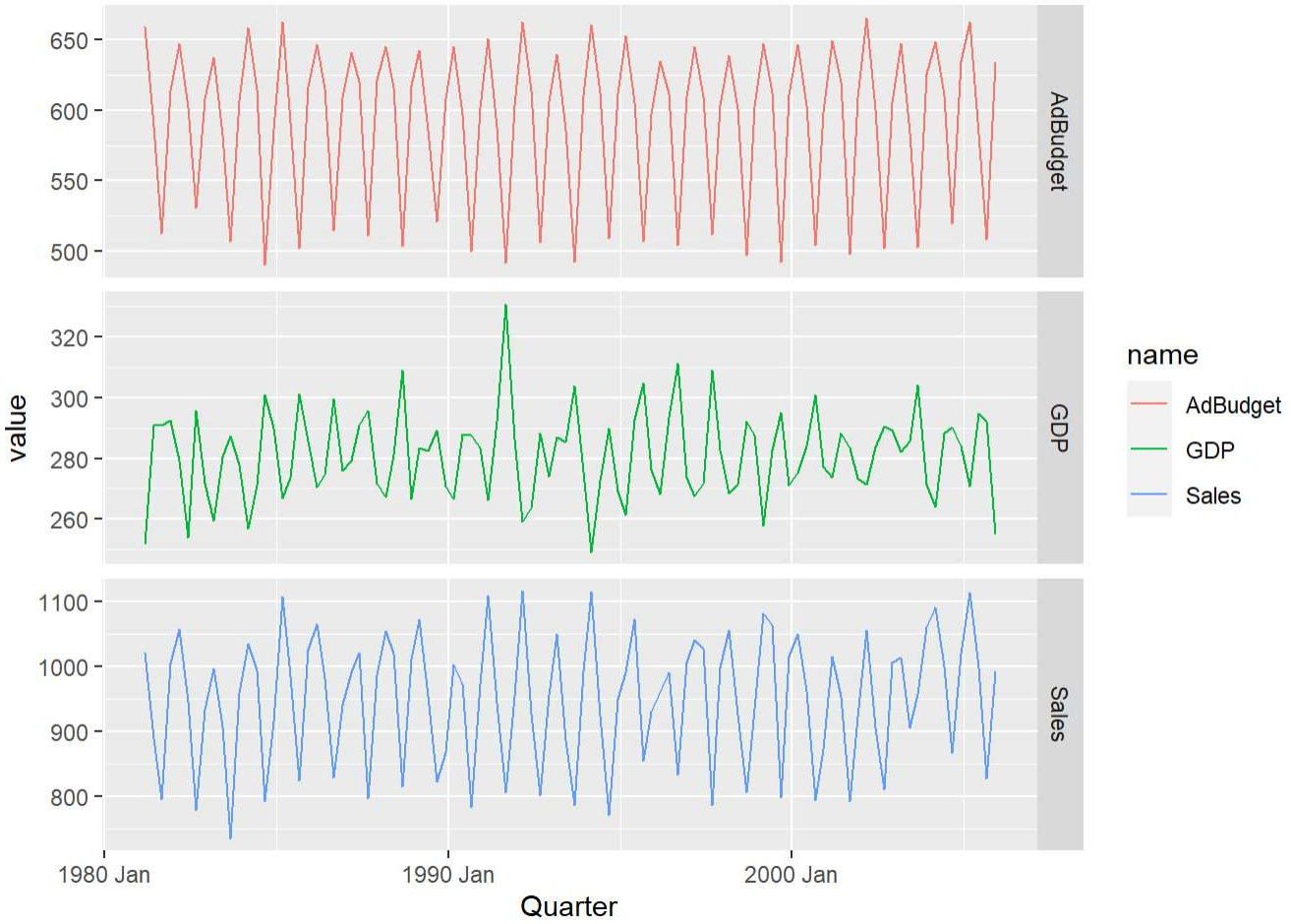
HIDE

```
#formatting as a tsibble
mytimeseries <- tute1 %>%
  mutate(Quarter = yearmonth(Quarter)) %>%
  as_tsibble(index = Quarter)
```

Construct time series plots of each of the three series

HIDE

```
mytimeseries %>%
  pivot_longer(-Quarter) %>%
  ggplot(aes(x = Quarter, y = value, colour = name)) +
  geom_line() +
  facet_grid(name ~ ., scales = "free_y")
```

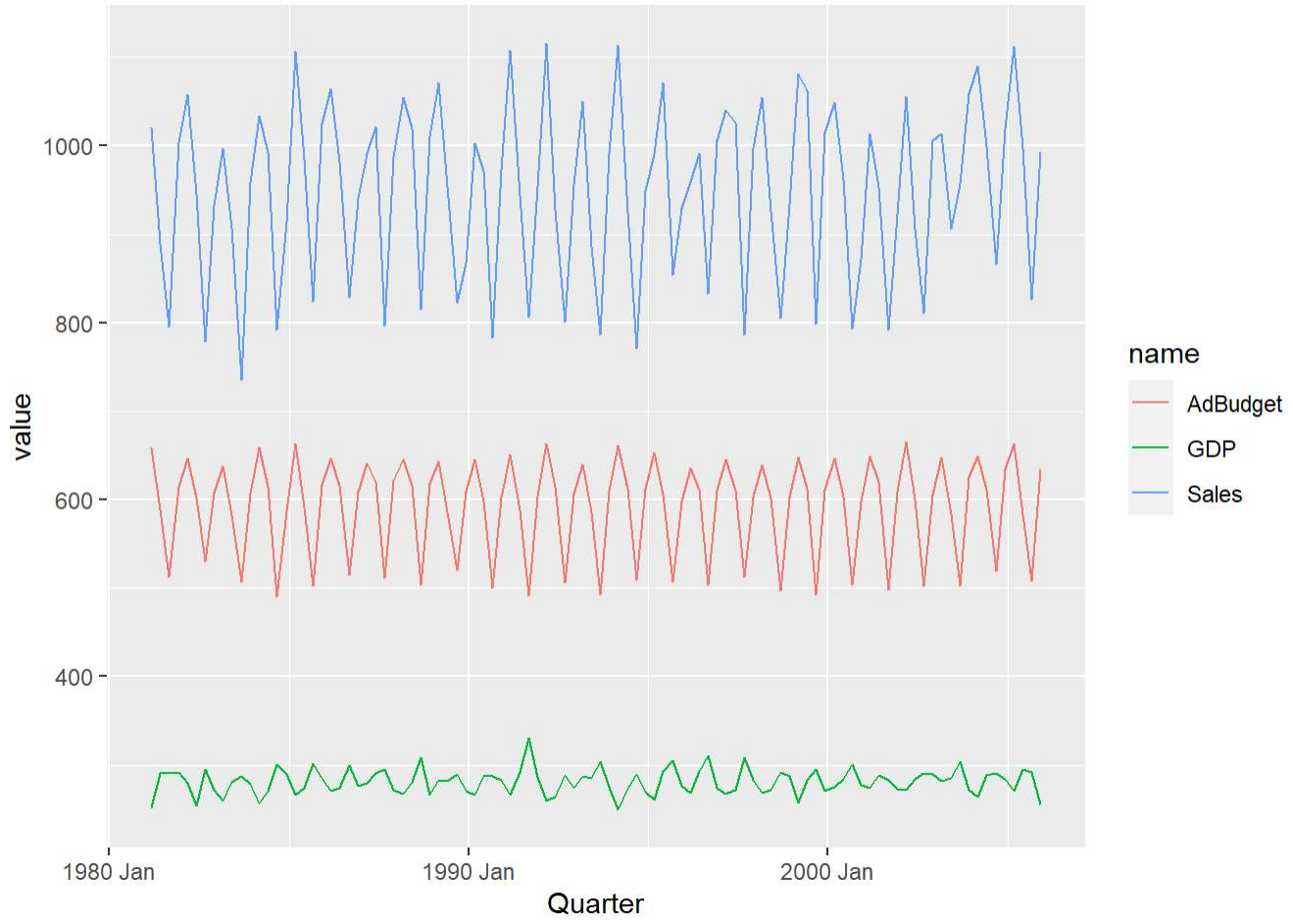


Check what happens when you don't include facet_grid().

Without `facet_grid()` each of the 4 variables are plotted on the same axes. In a different datasat the scale of each variable could have made these overlap each other, but the nature of each of these variables is that their y-axis scales don't appear to overlab since they are, of course, measured differently and measure different things.

HIDE

```
mytimeseries %>%
  pivot_longer(-Quarter) %>%
  ggplot(aes(x = Quarter, y = value, colour = name)) +
  geom_line()
```



Exercise 2.4

The USgas package contains data on the demand for natural gas in the US. Install the USgas package. Create a tsibble from us_total with year as the index and state as the key.

HIDE

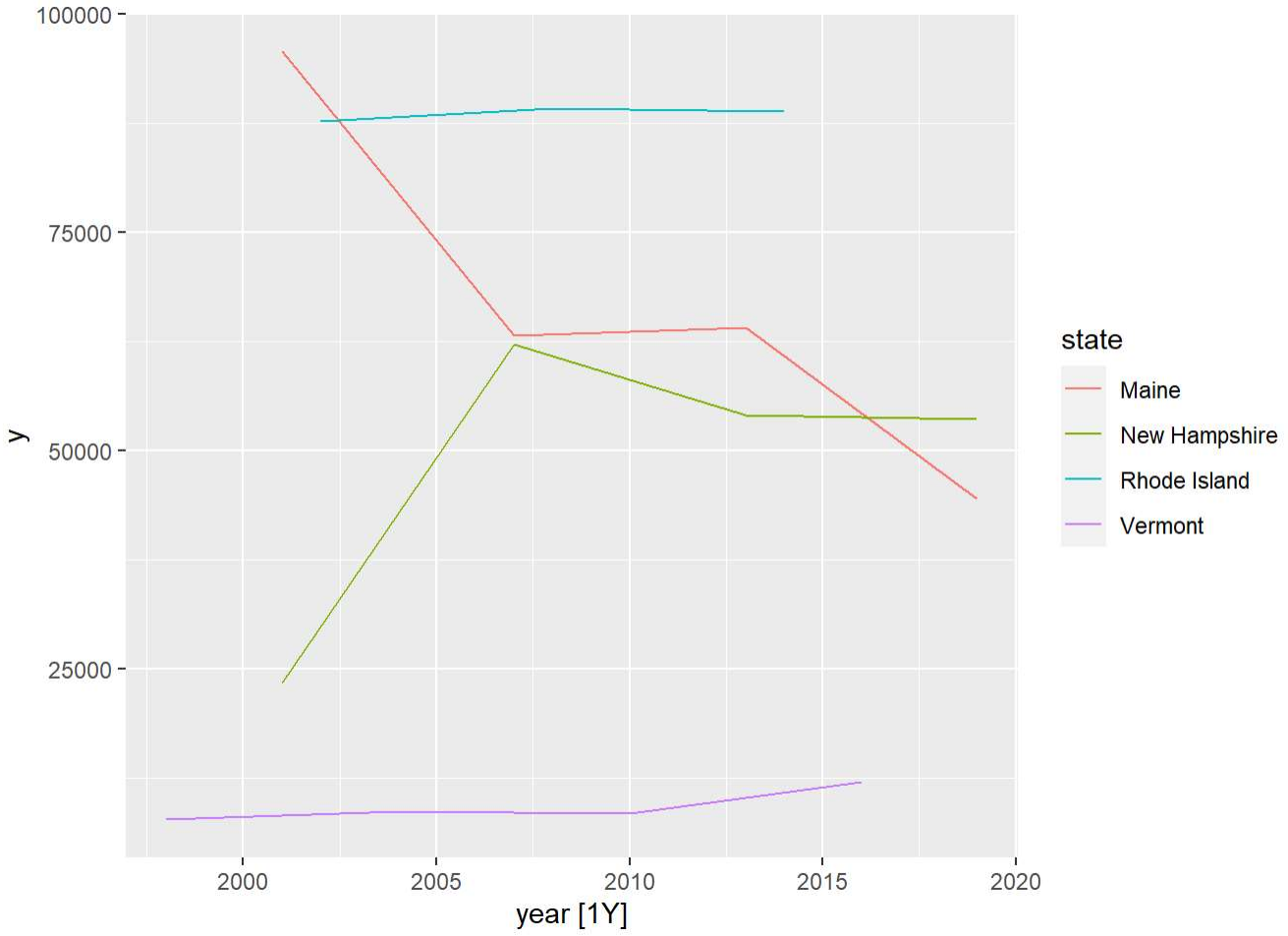
```
library("USgas")

usgas <- us_total %>%
  tsibble(
  key = state,
  index = year
)
```

Plot the annual natural gas consumption by state for the New England area (comprising the states of Maine, Vermont, New Hampshire, Massachusetts, Connecticut and Rhode Island).

HIDE

```
autoplot(usgas %>%
           filter(state == c("Maine", "Vermont", "New Hampshire", "massachusetts", "connet
icut", "Rhode Island")))
```



Exercise 2.5

Download tourism.xlsx from the book website and read it into R using readxl::read_excel(). Create a tsibble which is identical to the tourism tsibble from the tsibble package.

HIDE

```
tourism_raw <- readxl::read_excel("tourism.xlsx")

# make date class
tourism_raw$Quarter <- as.Date(tourism_raw$Quarter, "%Y-%m-%d")
# put in yyyy qq format
tourism_raw$Quarter <- yearquarter(tourism_raw$Quarter)

# put into tsibble
tourism_2 <- tourism_raw %>%
  tsibble(key = c(Region, State, Purpose, Trips),
  index = Quarter
)

# check tsibble from package to compare
#view(tourism)
```

Find what combination of Region and Purpose had the maximum number of overnight trips on average.

HIDE

```
# select needed variables, group by prompt, add trips_avg variable,
# drop Trips so you can de-dupe, then arrange by trips_avg in descending order
tourism_raw %>%
  select(c(Region, Purpose, Trips)) %>%
  group_by(Region, Purpose) %>%
  mutate(trips_avg = mean(Trips)) %>%
  select(-Trips) %>%
  distinct() %>%
  arrange(desc(trips_avg))
```

```
## # A tibble: 304 x 3
## # Groups:   Region, Purpose [304]
##    Region          Purpose  trips_avg
##    <chr>           <chr>        <dbl>
##  1 Sydney          Visiting      747.
##  2 Melbourne       Visiting      619.
##  3 Sydney          Business      602.
##  4 North Coast NSW Holiday       588.
##  5 Sydney          Holiday       550.
##  6 Gold Coast      Holiday       528.
##  7 Melbourne       Holiday       507.
##  8 South Coast     Holiday       495.
##  9 Brisbane        Visiting      493.
## 10 Melbourne       Business      478.
## # ... with 294 more rows
```

Create a new tsibble which combines the Purposes and Regions, and just has total trips by State.

I think I'm interpreting this prompt correctly,I've created a tsibble that maintains the Quarter variable as the index and for each quarter shows the total trips within that State with regard to the quarter and the purpose_region. Looking at the first few rows of data below, one could say that business trips to Canberra in the state of ACT decreased from Q1 to Q2 of 1998.

HIDE

```
tourism_states <- tourism_raw %>%
  select(c(Region, Purpose, Trips, State, Quarter)) %>%
  group_by(State, Quarter) %>%
  mutate(purpose_region = paste(Purpose, "-", Region),
         state_tot_trips = sum(Trips)) %>%
  select(-c(Region, Purpose, Trips)) %>%
  distinct() %>%
  as_tsibble(key = c(State, purpose_region),
             index = Quarter)

head(tourism_states)
```

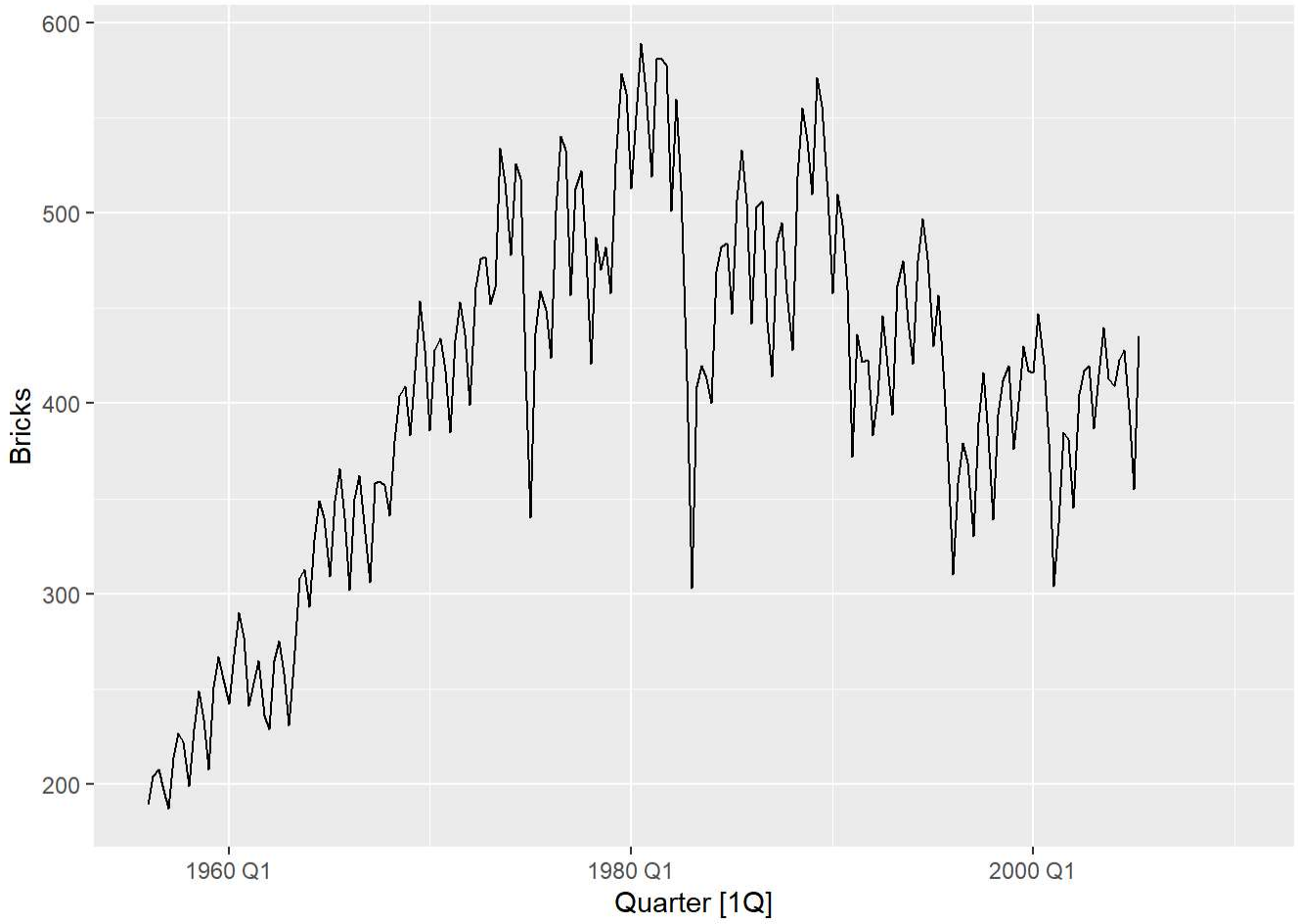
```
## # A tsibble: 6 x 4 [1Q]
## # Key:       State, purpose_region [1]
## # Groups:    State @ Quarter [6]
##   State Quarter purpose_region      state_tot_trips
##   <chr>   <qtr> <chr>                         <dbl>
## 1 ACT   1998 Q1 Business - Canberra            551.
## 2 ACT   1998 Q2 Business - Canberra            416.
## 3 ACT   1998 Q3 Business - Canberra            436.
## 4 ACT   1998 Q4 Business - Canberra            450.
## 5 ACT   1999 Q1 Business - Canberra            379.
## 6 ACT   1999 Q2 Business - Canberra            558.
```

Exercise 2.6

Create time plots of the following four time series: Bricks from aus_production, Lynx from pelt, Close from gafa_stock, Demand from vic_elec. Use ? (or help()) to find out about the data in each series.

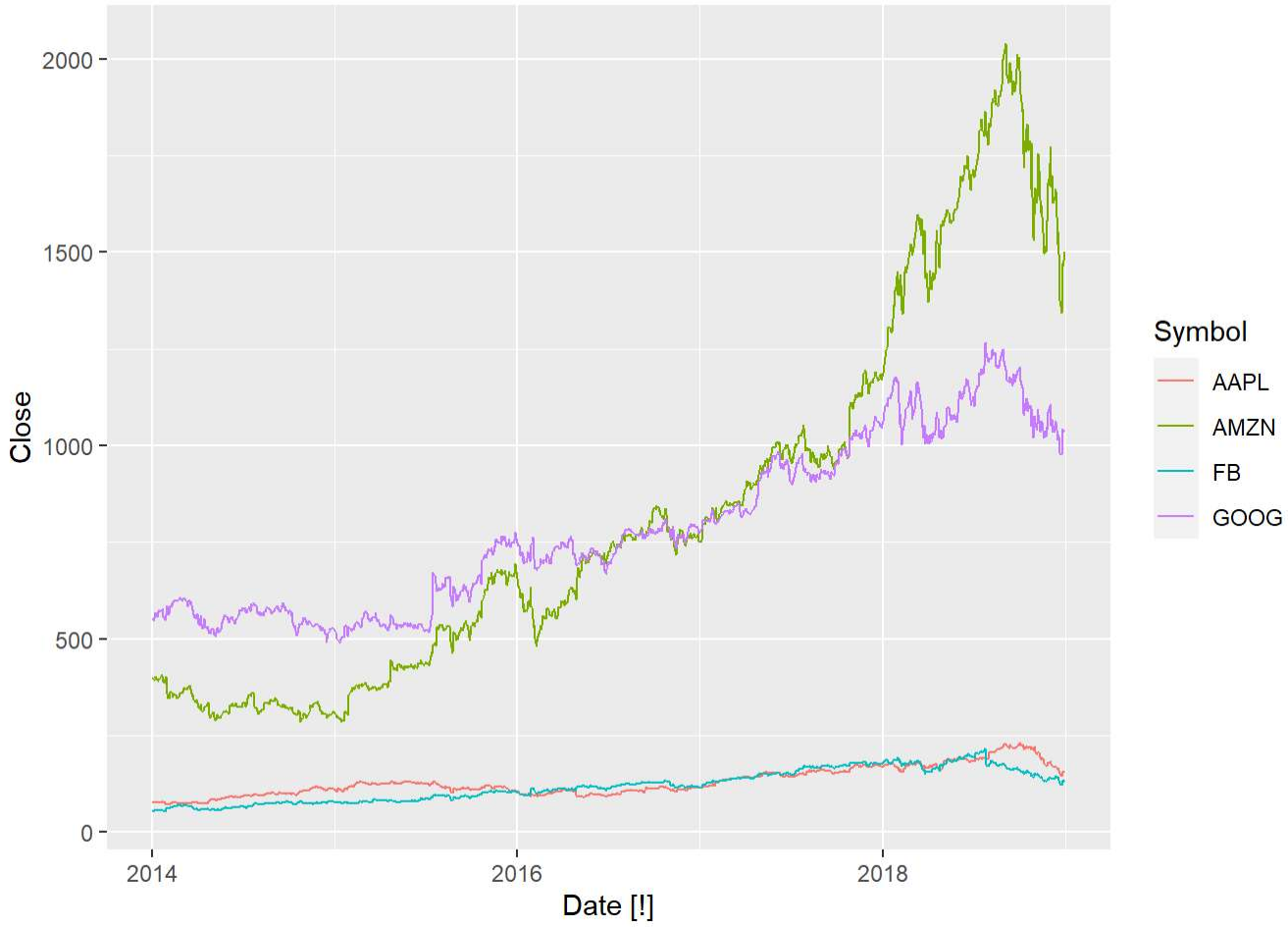
HIDE

```
autoplot(aus_production, Bricks)
```



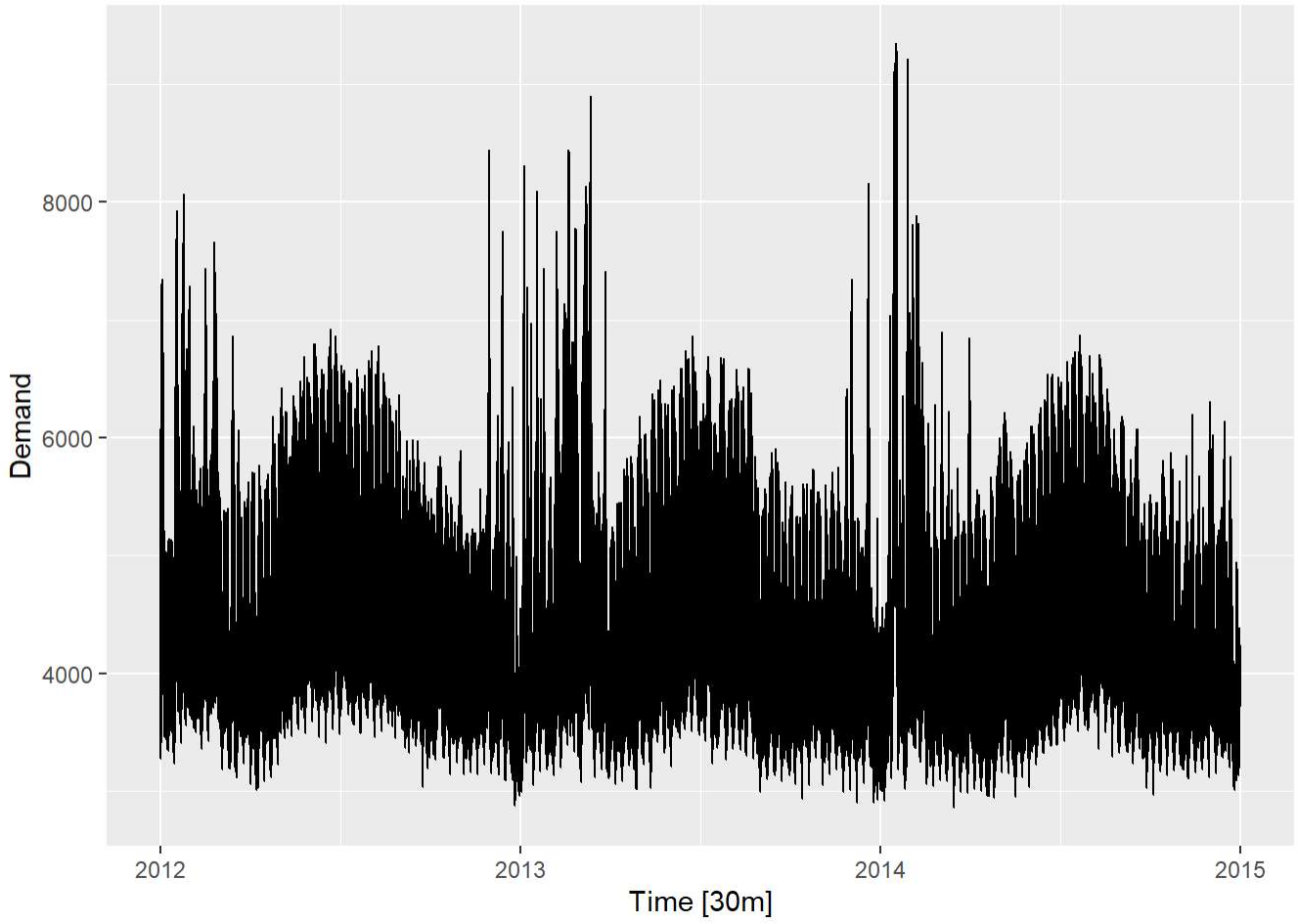
HIDE

```
autoplot(gafa_stock, Close)
```



HIDE

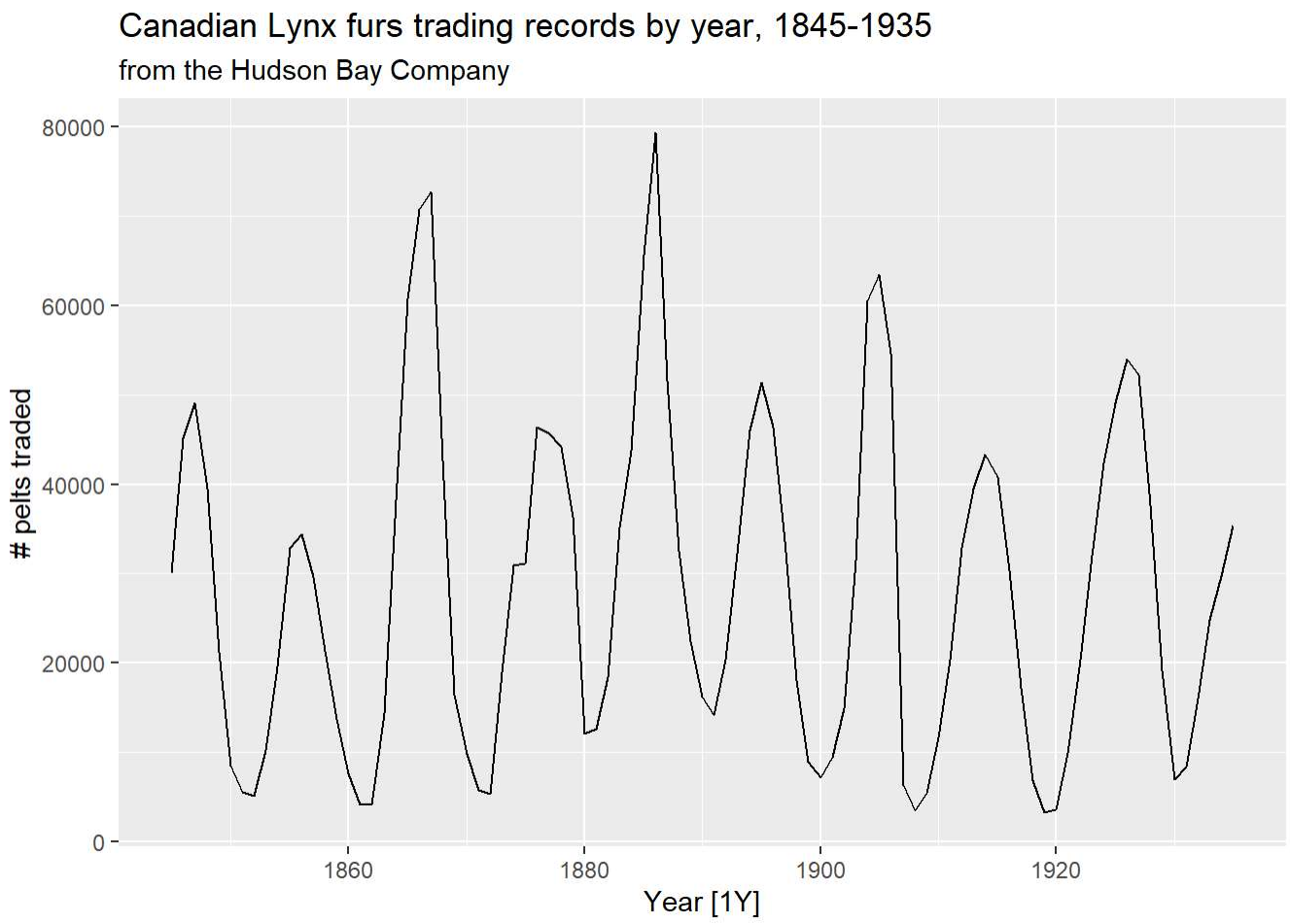
```
autoplot(vic_elec, Demand)
```



For the last plot, modify the axis labels and title.

HIDE

```
autoplot(pelt, Lynx) +
  labs(title = "Canadian Lynx furs trading records by year, 1845-1935",
       subtitle = "from the Hudson Bay Company",
       y = "# pelts traded")
```



Exercise 2.8

Monthly Australian retail data is provided in aus_retail. Select one of the time series as follows (but choose your own seed value):

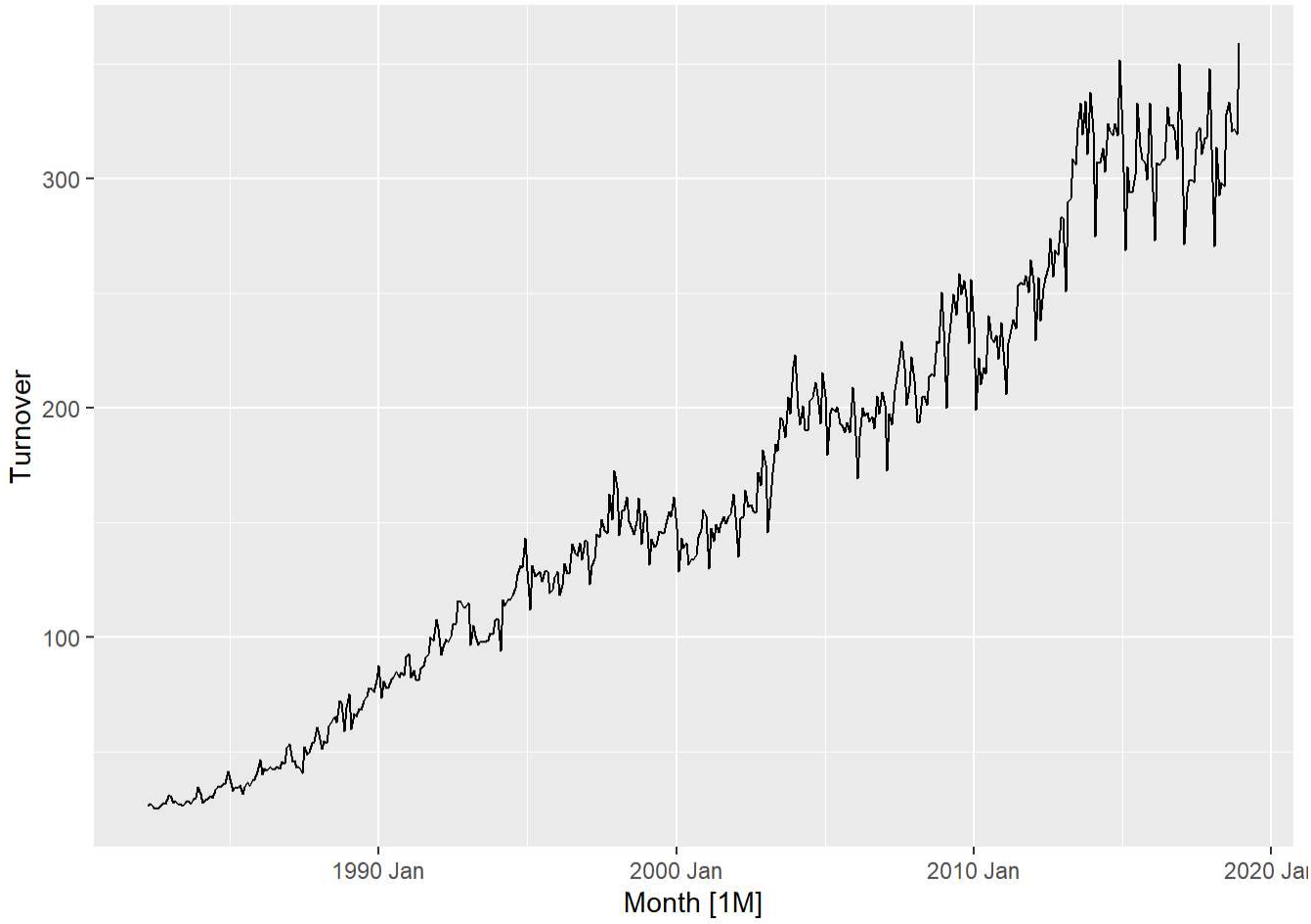
HIDE

```
set.seed(8675309)
myseries <- aus_retail %>%
  filter(`Series ID` == sample(aus_retail$`Series ID`,1))
```

Explore your chosen retail time series using the following functions:autoplot(), gg_season(), gg_subseries(), gg_lag(), ACF()

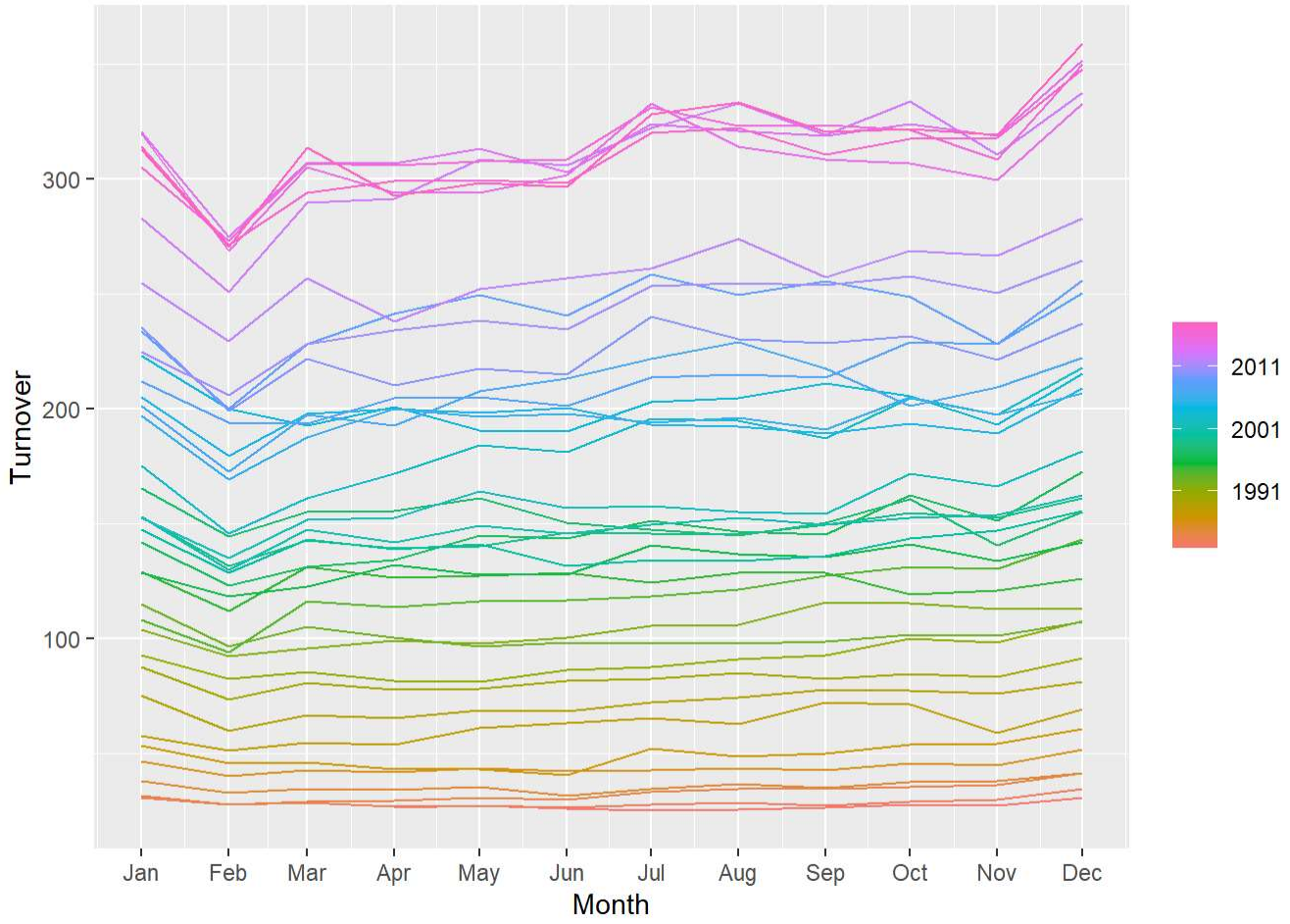
HIDE

```
autoplot(myseries)
```



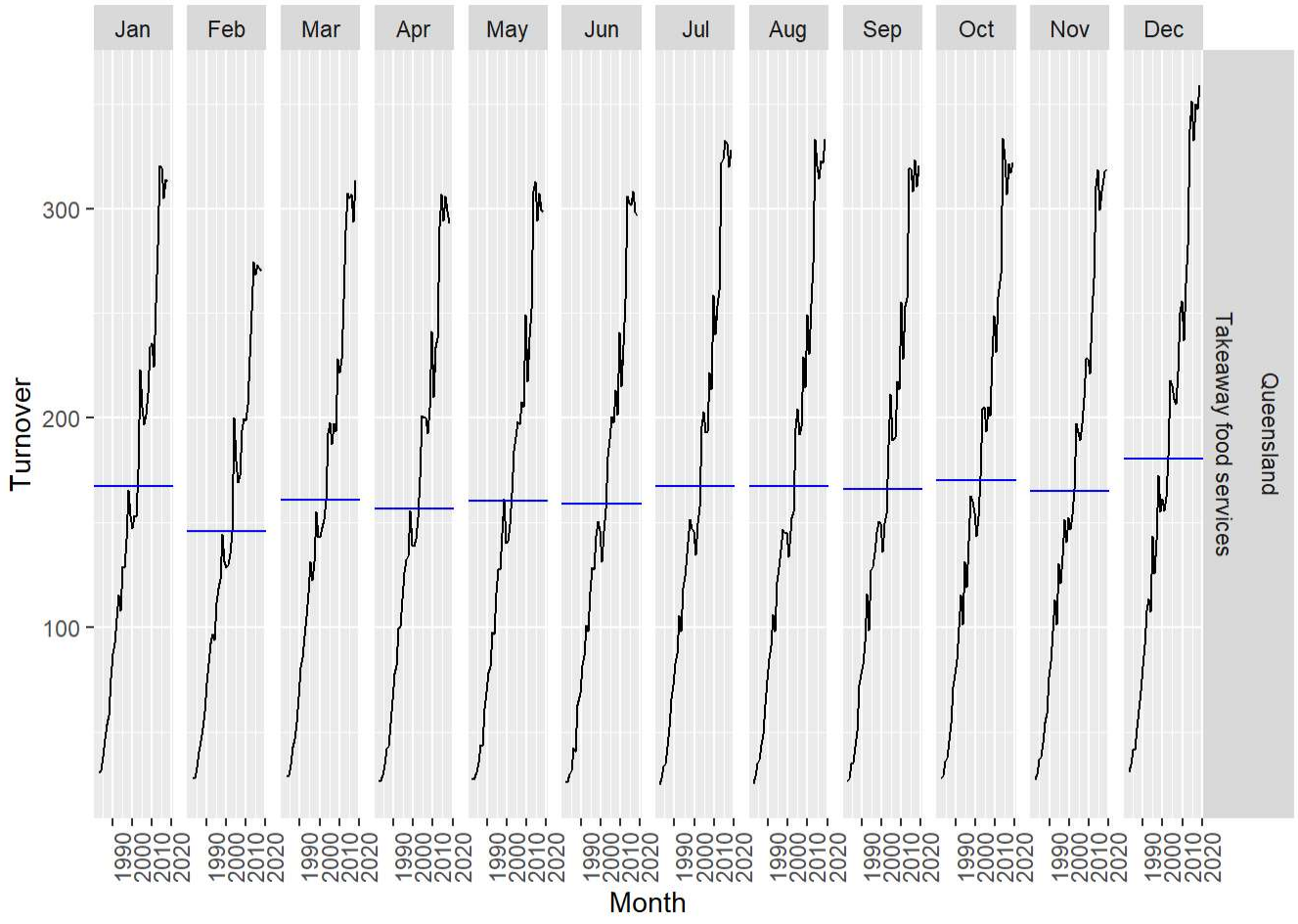
HIDE

```
gg_season(myseries)
```



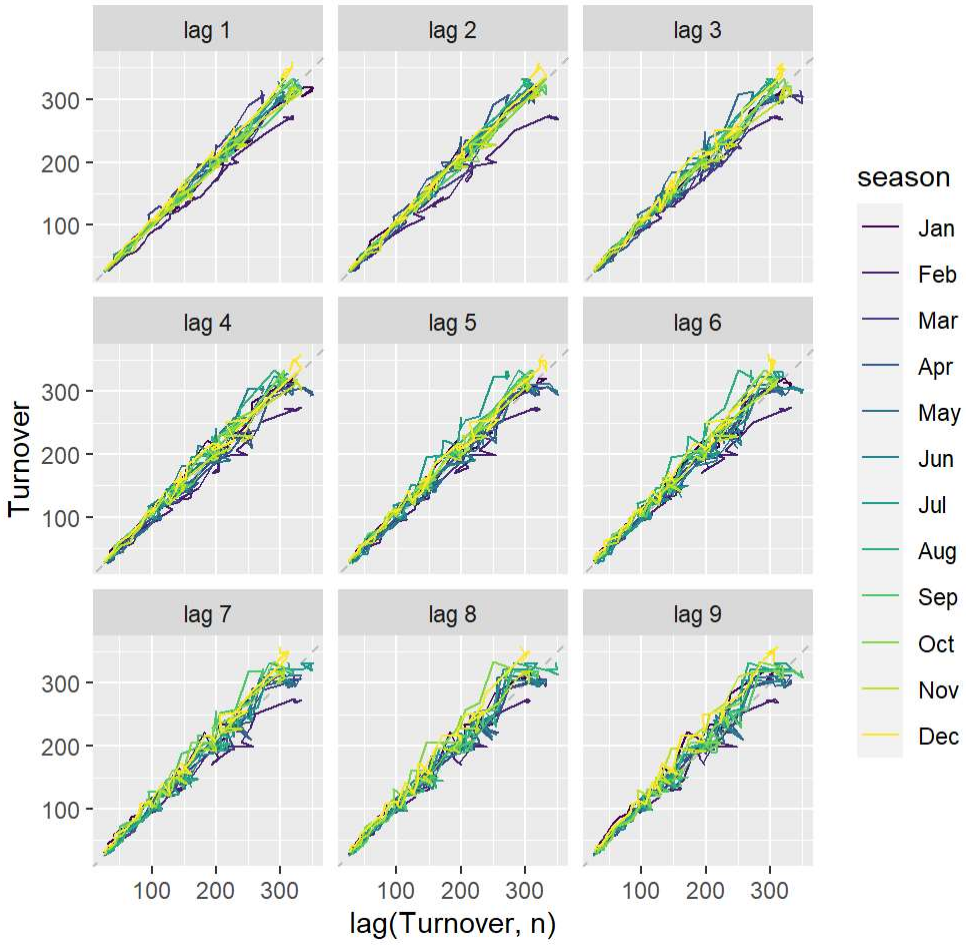
HIDE

```
gg_subseries(myseries)
```



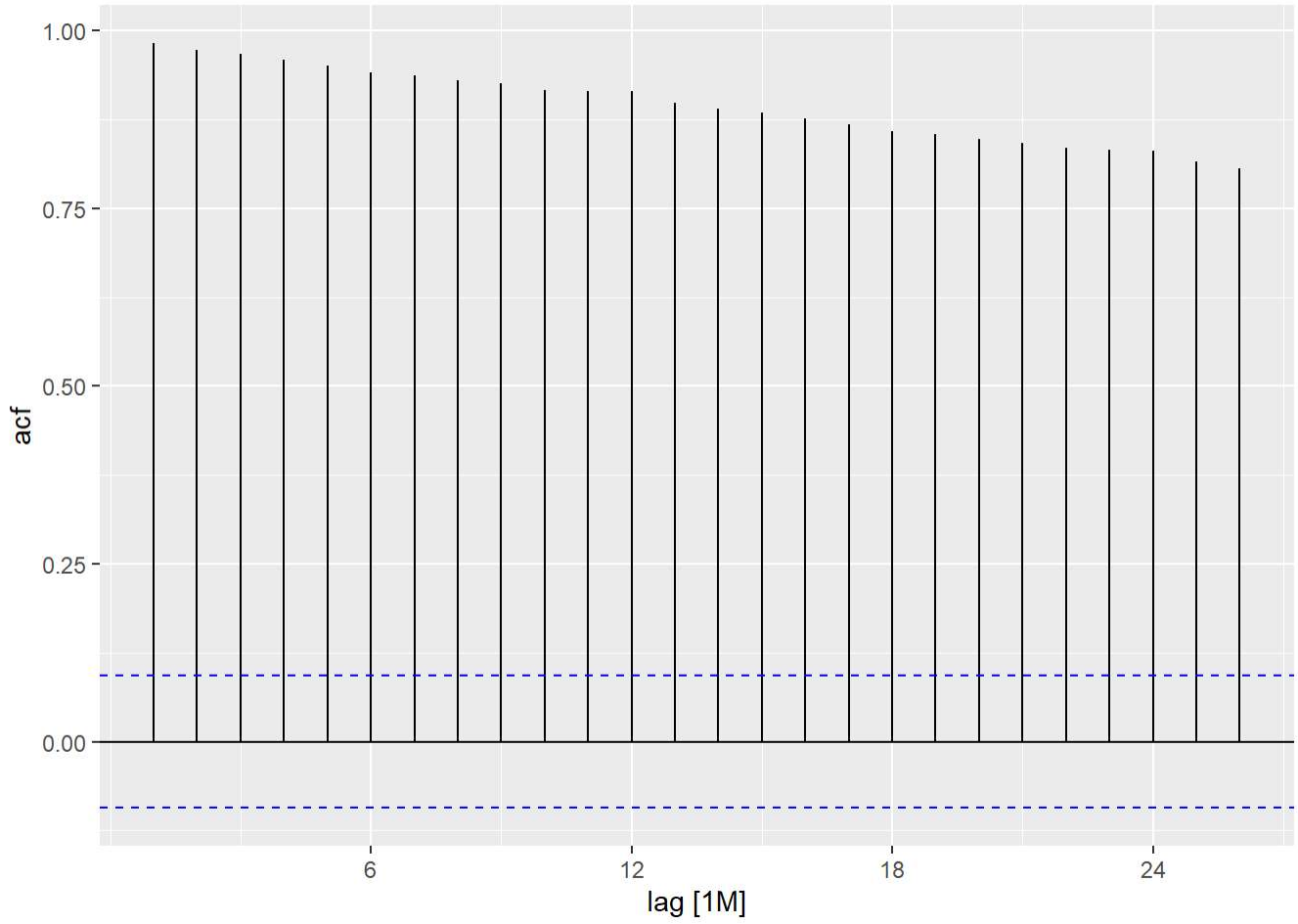
HIDE

```
gg_lag(myseries)
```



HIDE

```
myseries %>%
    ACF(Turnover) %>%
    autoplot()
```



Can you spot any seasonality, cyclicity and trend? What do you learn about the series?

The autoplot() graph shows a definite upward trend in retail turnover, with a hint of some sort of repeated pattern but it's hard to tell from this generic plot. In looking at the gg_season() plot we again see more recent years (pink and purple) higher on the y-axis of retail turnover. Further, there is appears to be a dip in many years in February, with an increase in December. The gg_subseries() plot is a bit hard to read with the year-span so thin, but seeing the blue lines of what I assume is a median does show the dip in February and the slight increase in December I noticed earlier.

The gg_lag() plot shows a high correlation across all lags, while the ACF and plot might show a very slight scalloped shape which hints at the upward trend and minimal seasonality playing off of each other in this chart.

I'd feel fairly safe, at this stage of exploration, saying there appears to be an upward trend of retail turnover increasing as well as consistent peaks in December, dropping in January, and dropping to it's lowest in February. March-November appear to be relatively steady.